STAT 515 hw 10

Two-sample testing, comparative experiments and ANOVA

Attach a sheet with any R plots and R code printed on it. You may write out your other answers by hand if you want. Just try to make it easy for me grade!!

1. In a study of how different types of greetings transmit bacteria, a sterile glove was donned, dipped in bacteria, and then used in a handshake, a high five, or a fist bump with a hand wearing a sterile glove. Afterwards the bacteria on the sterile glove were counted. These data come from exercise 9.24 of McClave and Sincich (2016). Read the data into R using

```
handshake <- c(131,74,129,96,92)
highfive <- c(44,70,69,43,53)
fistbump <- c(15,14,21,29,21)
```

   It is of interest to study differences among the mean bacteria counts expected from these types of greeting, which we may denote by $\mu_{\text{handshake}}$, μ_{highfive}, and μ_{fistbump}.

   (a) Use R to get a 99% confidence interval for $\mu_{\text{handshake}} - \mu_{\text{highfive}}$ assuming $\sigma^2_{\text{handshake}} = \sigma^2_{\text{highfive}}$. Use the command

```
t.test(handshake,highfive,conf.level=.99,var.equal=TRUE)
```

   (b) Use R to get a 90% confidence interval for $\mu_{\text{highfive}} - \mu_{\text{fistbump}}$ under the assumption that $\sigma^2_{\text{highfive}} \neq \sigma^2_{\text{fistbump}}$.

   (c) Use the command

```
boxplot(handshake,highfive,fistbump)
```

   to get boxplots of the data. Turn in this plot.

   (d) Based on the boxplots, comment on whether you should assume $\sigma^2_{\text{highfive}} = \sigma^2_{\text{fistbump}}$.

   (e) Use R to test the hypotheses

$$H_0\colon \mu_{\text{handshake}} - \mu_{\text{fistbump}} = 0 \text{ versus } H_1\colon \mu_{\text{handshake}} - \mu_{\text{fistbump}} \neq 0$$

   at the $\alpha = .05$ significance level. You must decide whether to put `var.equal=TRUE` or `var.equal=FALSE`. Say whether you reject H_0 and why based on the output.

   (f) Suppose an investigator wanted to do an ANOVA for these data, where `handshake`, `highfive`, and `fistbump` are considered treatments. Which one of the ANOVA assumptions does not appear to be satisfied for these data?

2. Execute the commands below in R to read in some data. The data points are the number of crashes (average per year) due to drivers' running red lights at thirteen intersections before and after the installation of red light cameras. These data come from exercise 9.53 of McClave and Sincich (2016). Read the data into R with the commands

```
before <- c(3.6,.27,.29,4.55,2.6,2.29,2.4,0.73,3.15,3.21,.88,1.35,7.35)
after <- c(1.36,0,0,1.79,2.04,3.14,2.72,0.24,1.57,0.43,0.28,1.09,4.92)
```

It is of interest to see whether the installation of a camera reduces the number of crashes due to running red lights.

(a) Compute the differences in the numbers of crashes at the intersections:

```
diff <- before - after
```

Give the mean before-minus-after difference from the sample.

(b) Formulate a set of hypotheses for testing whether the installation of cameras is effective in reducing the number of accidents. Use μ_{diff} to denote the mean difference. Hint: This is not a two-sample problem but a one-sample problem, even though it may look like a two-sample problem because two sets of data have been given. Ask yourself, if the cameras are effective in reducing the number of crashes, should μ_{diff} be greater than or less than zero?

(c) Create a Normal quantile-quantile plot of the differences and comment on whether you think the differences are Normally distributed.

(d) The `t.test()` function in R can be used in the one-sample setting too. Use the command

```
t.test(diff,alternative=" ")
```

to get a p-value for the test. You must decide whether to put `greater`, `less`, or `two.sided` in for the alternative.

(e) What is your decision at the $\alpha = .05$ significance level? Are the cameras effective?

3. It is of interest whether soil scouring has any effect on whether a tree growing in a flood plain falls. Researchers subjected trees to three different degrees of soil scouring (none, shallow, and deep) and then measured the maximum resistive bending moment of the tree trunk bases. Read the data, which comes from exercise 10.36 of McClave and Sincich (2016), into R in preparation to run an ANOVA using the commands

```
maxresist <- c(23.68,8.88,7.52,25.89,22.58,11.13,29.19,
        13.66,20.47,23.24,4.27,2.36,8.48,12.09,3.46)
soilsc <- c( rep("none",5),rep("shallow",5),rep("deep",5))
```

(a) If it is of interest whether soil scouring has any effect on the mean maximum resistive bending moment of the tree trunk bases, what are the relevant null and alternate hypotheses in terms of μ_{none}, μ_{shallow}, and μ_{deep}?

(b) Execute the command

```
plot(lm(maxresist~soilsc))
```

and press enter in the console to scroll through four different plots. One of them is a Normal quantile quantile plot of the residuals. Turn in this plot and comment on whether you think the residuals are Normal.

(c) Execute `plot(lm(maxresist ~ soilsc))` again and look at the Residuals vs Fitted plot. Turn in this plot and comment on whether you think the variance of the response is the same in all three treatment groups.

(d) Enter the command

```
anova(lm(maxresist~soilsc))
```

to get the ANOVA table. Turn in this table.

(e) If the ANOVA assumptions are satisfied, what do you conclude about the effect of soil scouring at the $\alpha = 0.05$ significance level?

4. It is of interest whether the temperature has any effect on the mean ethanol concentration in bio-fuel produced in a fermentation process. An experiment is run under the temperatures 30°, 35°, 40°, and 45° degrees Celsius. Read the data, which come from exercise 10.39 of McClave and Sincich (2016), into R in preparation for ANOVA, with the commands

```
ethanol <- c( 103.3,103.4,101.0,101.7,102.0,101.1,97.2,96.9,96.2,55.0,56.4,54.9)
temp <- c(rep("30deg",3), rep("35deg",3),rep("40deg",3),rep("45deg",3))
temp <- as.factor(temp)
```

(a) If it is of interest whether the temperature has any effect on the mean ethanol concentration, what are the relevant hypotheses in terms of $\mu_{30°}$, $\mu_{35°}$, $\mu_{40°}$, and $\mu_{45°}$?

(b) Do part (b) of Question 3 for the ethanol data.

(c) Do part (c) of Question 3 for the ethanol data.

(d) Do part (d) of Question 3 for the ethanol data.

(e) If the ANOVA assumptions are satisfied, what do you conclude about the effect of temperature on the ethanol concentration at the $\alpha = 0.01$ significance level?

References

McClave, J. and Sincich, T. (2016). *Statistics*. Pearson Education.